

博士論文審査報告書

モンテカルロ法に基づく
ベイズ学習の拡張に関する研究

申 請 者

中田	洋平
Nakada	Yohei

電気・情報生命専攻
学習型信号・情報処理システム研究

2009 年 2 月

画像・文書・生物などを含め、広範な学術領域のデータベースの飛躍的拡充と共に、高性能計算機の普及、インターネットを始めとする情報ネットワーク網の整備など、社会全体の情報インフラの革命的な発展が 20 世紀終盤に訪れた。その結果、現代社会は高度情報化社会へと変革を遂げ、個人、企業、官公庁、教育機関、研究機関などあらゆる主体が膨大な量のデータを蓄積するに至った、また、その主体間を様々な種類のデータが情報ネットワーク網を介して電子的に流通するようにもなった。そして、蓄積・流通するデータの形式は多岐に渡る。例えば、個人が接する機会の多い文書情報、画像情報などや、企業の保持する顧客情報、科学分野に関連する気象情報、遺伝子情報を含む生物情報などもある。このような中で、これまで得られてきた情報科学や統計科学、または、その周辺分野の研究成果を活用し、膨大かつ多様なデータから有用な“何か”を自動的に獲得する技術群も急速な発展を見せてきている。例えば、データから有用な知識を抽出することを目指すデータマイニング技術、データから特定のパターンを見出すようなパターン認識技術、データからその背後にある母集団の性質を推定するような統計技術、または、データを用いて予測機能などを持つ機械を構築するような機械学習技術などである。情報化社会の更なる高度化の要請もあり、これらの技術群の発展が引き続き望まれている。このような中、現在、前述のような技術群に対する実践性の向上策として、先験知識を推論体系に取り込む柔軟性があり、比較的高い頑健性を持つベイズ推論（Bayesian inference）が注目を浴びてきている。

申請者の研究は、このようなベイズ推論の機械学習分野への導入、いわゆるベイズ学習に関するものであり、本論文は申請者が遂行してきたベイズ学習の拡張に関する研究を纏めたものである。本論文は、研究成果の位置付けを明示化するためにベイズ学習の全体像を申請者独自の視点で再整理した第 1 部と、申請者の成した具体的な研究群を纏めた第 2 部とで構成されている。以降、本論文の記載内容を第 1 部、第 2 部ごとに具体的に評価していく。

本論文の第 1 部では第 1 章～第 3 章に渡り、大局的な観点に基づいてベイズ学習の輪郭を頭にしている。第 1 部の構成をより具体的に見れば、第 1 章では、ベイズ学習の大きな意味での包括分野であるデータ分析分野全体の現況や、ベイズ学習の基となるベイズ推論や機械学習について説明している。また、その途中でベイズ推論と密接に関連するものの、若干異なる統計学的枠組みである近代統計学についても概説している。続く第 2 章では、ベイズ学習の枠組みに関連する事項をベイズ学習の類型なども含めて述べている。最後の第 3 章では、ベイズ学習を実際に用いる際の実装に関連する内容について説明している。この中では数値実験例を用いて幾つかの実装手段の比較も実施している。このように第 1 部では、申請者の成した研究の主題である

ベイズ学習の拡張に関わる成果を、申請者の視点を保持しつつ実証的に明示化しようとしている点が高く評価できる。その記載内容も非常に多くの文献を引用しながら多角的に検証しており、信頼性や妥当性も申し分ない内容と言える。特に第2部で述べる研究群につながるベイズ学習のこれまでの発展について、何が要因としてあるかを深い洞察とともに論じられている点は評価に値する。その洞察の一例を述べたい。ベイズ学習では事前分布と呼ばれるパラメータの確率分布を設定する必要がある。このことは、前述のような先験知識を推論に組み込むことを可能とする一方で、推論結果がその事前分布の設定に依存することが問題視されることもあった。この点が長い間ベイズ学習の応用可能性を限定してきた。しかしこの事前分布設定の問題を緩和する階層ベイズ (hierarchical Bayes) や経験ベイズ (empirical Bayes) と呼ばれる枠組が発展してきたことにより、ベイズ学習の枠組の中で先験情報を推論に組み込む度合いなどの調整が可能となり、更にこれがベイズ学習の応用可能性を大きく広げることに寄与したと言ってよい。本論文では、このような階層ベイズや経験ベイズについても適切な洞察とともに論じており、更に多くの他の事前分布の研究についても整理している。また、ベイズ推論の過程に存在する確率的な積分計算も、事前分布設定の問題と同様にベイズ推論の応用可能性を限定する大きな要因であったと言えるが、近年、確率積分計算に対する実装手段 (近似解法) が確立されてきたことに伴い、従来の統計モデルの範囲を超え機械学習技術で用いられるモデルなども含めた広い範囲でその応用可能性が向上してきている。本論文ではこのことも的確に記載している。更にはベイズ推論の応用可能性を向上に導いたこのような実装手段を3つの分類方式にてその実装手段を整理し、その性能比較についても数値実験例を用いて具体的に表している。本論文の第1部は、このようなベイズ学習への近年の発展について適切に記載されており、後述の第2部に示す研究群の位置付けを明確化する内容と言える。

第2部では、第1部でその位置付けが述べられたベイズ学習に対して、申請者が実施してきた具体的な研究が第4章～第7章に渡り論じられている。第2部の構成は以下のようになっている。まず、第4章では、一括型ベイズ学習の時系列予測への拡張を論じている。本章では、特にデータの背後にある系が非線形性を持ち、更にその形式が未知の場合の非線形時系列予測問題を対象としている。更に、多層パーセプトロンなどの非線形機械学習モデルを用いて時系列データの背後にある系を学習する方法や、ベイズ学習から得られたパラメータのサンプルを用いて将来の時系列を予測する方法を提案している。これにより、従来技術と比べより適切な予測分布の近似や、従来は困難であった予測時のエラーバーの算出などを可能とし、実際に予測困難性の高いカオス力学系から生成した時系列データによる数値実験例と実データとして空調機熱負荷予測時系列データを用いた例を用いて提案手法の高い性

能を示している．第5章では，逐次型ベイズ学習の変化検出への拡張を論じている．本章では，特に，データの背後に潜む系が非線形性を持ち，かつ，その形式が未知の場合を対象としている．このような対象に対して，本章では，逐次モンテカルロ法に基づく逐次型ベイズ学習を拡張した逐次型変化検出手法を提案している．更に回帰問題に基づく数値実験例とカオス時系列予測問題に基づく数値実験例とを用いた検証実験と ROC 分析を実施し提案手法の高い検出性能を示している．第6章では，これまでに無い逐次型ベイズ学習用の事前分布に関する研究を記載している．より適切な事前分布が求められていたにも拘らず，これまでは逐次型ベイズ学習に対する事前分布についてはあまり研究が成されていなかった．申請者は，従来とは異なる観点を導入し設計した「自然逐次事前分布」を提案している．これは，従来の事前分布に比べ，数理モデルの時間的变化をより適切に考慮することを可能とするものであるとし，時間的に変化のある非線形回帰問題に基づく数値実験例を用いてその学習性能を示している．更に，この自然逐次事前分布について数理的観点からの考察も行い，幾つかの最適性について明らかにしている．第7章では，集団学習アルゴリズムの1つであり，高精度なことで注目を集めるブースト学習に関する成果を報告している．より具体的には，ブースト学習にベイズ学習の要素を導入した新たなブースト学習「ギブス・ブースト」を提案している．誤ラベルの混入を考慮した2値判別問題に基づく数値実験例を用いて，提案手法を検証しその判別能力を示している．また，提案手法の数理的性質を明らかにする目的で，提案手法で用いるモデルのある種の極限での振る舞いを考察し，他のベイズ学習との関係を明示化している．このように，本論文の第2部では幾つかのベイズ学習の拡張を提案しており，それらが従来技術を上回る性能を示している点が評価できる．その幾つかについては，本論文内で適正な考察により数理的な性質も明らかにされている点も高く評価できる．

以上のように第1部，第2部ともに高い水準を保持した内容であり，ベイズ学習への貢献は高く評価すべきであると考えられる．よって本論文は博士（工学）の学位論文として価値あるものと認める．

2009 年 1 月

審査員

（主査）	早稲田大学	教授	工学博士	（早稲田大学）	松本	隆
	早稲田大学	教授	工学博士	（早稲田大学）	内田	健康
	早稲田大学	教授	博士（工学）	（東京大学）	村田	昇
	早稲田大学	教授	博士（工学）	（早稲田大学）	渡邊	亮
	早稲田大学	准教授	医学博士	（京都大学）	井上	真郷